DOCUMENT RESUME

ED 285 886 TM 870 461

AUTHOR Engelhard, George, Jr.
TITLE The Use of Opportunity To Learn To Obtain Curriculum-Based Estimates of Student Achievement.
PUB DATE Apr 87
NOTE 36p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Washington, DC, April 21-23, 1987).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Achievement Tests; Individual Testing; *Mathematics Achievement; Mathematics Tests; Maximum Likelihood Statistics; National Surveys; Scoring; Secondary Education; *Secondary School Mathematics; Test Bias; *Test Validity
IDENTIFIERS *Curriculum Related Testing; Second International Mathematics Study

ABSTRACT

A method for obtaining curriculum-based estimates of student achievement is described. These estimates can be obtained through the use of a set of curriculum weights designed to reflect student opportunity to learn. If the match between the presented curriculum and test content is low, then the curriculum weights can be used to minimize potential curriculum bias which may occur when the students have not had an opportunity to learn the objectives measured by the test items. A small example and application of the curriculum-based estimates using a 40-item arithmetic test from the Second International Mathematics Study was used to illustrate the method. Posttest data from 2,606 eighth and twelfth grade students in 165 U.S. classrooms were included in the final sample. It is concluded that curriculum-based estimates offer a practicable approach for individualized scoring based on content considerations. (Author/GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED285886

The Use of Opportunity to Learn to Obtain Curriculum-Based Estimates of Student Achievement

George Engelhard, Jr.

Emory University

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

G. Engelhard, Jr.

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

RUNNING HEAD - Curriculum-Based Estimates

Paper presented at the annual meeting of the National Council on Measurement in Education in Washington, DC, April 1987.

TM 870 461

Abstract

A method for obtaining curriculum-based estimates of student achievement is described. These estimates can be obtained through the use of a set of curriculum weights designed to reflect student opportunity to learn. If the match between the curriculum and test content is low, then the curriculum weights can be used to minimize potential curriculum bias which may occur when the students have not had an opportunity to learn the objectives measured by the test items. A small example and application of the curriculum-based estimates using arithmetic items from the Second International Mathematics Study (N = 510) was used to illustrate the method. Curriculum-based estimates offer a practicable approach for individualized scoring based on content considerations.

KEY WORDS: Customized tests, opportunity to learn, individualized test scoring, item response theory, curriculum bias, mathematics achievement

The Use of Opportunity to Learn to Obtain Curriculum-based Estimates of Student Achievement

One of the advantages of item response theory over classical test theory is that it becomes possible in principle to tailor a unique set of appropriate items for each student (Hambleton & Swaminathan, 1985). One application of this is the use of computers to administer adaptive tests which are individually tailored for each student (Weiss, 1982). In computerized adaptive testing situations, item "appropriateness" is defined primarily in terms of item difficulty; each student is presented with a tailored set of items selected to be in the appropriate range of difficulty given an initial estimate of student achievement. The key idea is that by administering appropriate items, errors in the measurement of the student due to guessing and carelessness will be minimized.

Is the difficulty of the item the only criterion that should be used to individualize tests in order to select, administer and score an "appropriate" set of items? In this paper, I plan to argue that within the context of school achievement testing it is also important to consider the potential effects of the school curriculum on the test scores. Test items can be tailored on the basis of content considerations which reflect a student's opportunity to learn rather than simply on item difficulty. I will describe and illustrate an approach which can be used for obtaining curriculum-based estimates of student achievement. This approach can be used to determine whether or not the curriculum has a significant impact on the estimates of student achievement.

Background

The estimation of student achievement always includes a certain amount of error. In achievement testing, one major source of error which has been given considerable attention is guessing. Since the introduction of multiple-choice test items, the problem of guessing has been recognized. This source of error is generally a random source of error in test scores. Another important source of error in achievement test scores which has not received as much attention is curriculum bias. A curriculum bias may occur when there is a lack of overlap between the objectives measured by the test items and the objectives which the students have had an opportunity to learn. For example, if the students have not had an opportunity to learn about derivatives in their calculus class, then the probability of succeeding on test items reflecting this objective will be decreased. To the extent that the learning of a set of objectives is dependent on the school curriculum, the decrease in student test scores may be considered a curriculum bias. The key idea here is that the degree of overlap between what is covered in the curriculum and what is tested may be introducing a systematic error or bias into the estimates of student achievement. Curriculum bias reflects the difference in the estimates of student achievement between the obtained score and the "true" score that the student might have obtained if he or she had had the opportunity to learn the objectives measured by the test items.

Is this potential curriculum bias significant? There has been some disagreement in the literature about the effects of lack of overlap between what is tested and what is taught. The views range

from Mehrens and Phillips (1986) who concluded that "neither curricular match judged by district personnel or textbook series used had a significant impact on standardized test scores" (p. 185) to the view of Pelgrum, Eggen and Plomp (1986) that opportunity to learn was an important variable in their study of the implemented and attained mathematics curriculum in eighteen countries. Several other studies provide support for the importance of overlap (Anderson, 1985; Borg, 1979; Jenkins and Pany, 1978; Miller, 1986). Perhaps the best way to answer this question is to view the significance of curriculum bias as being dependent on the testing situation. Whether or not curriculum bias is significant is an empirical question which should be explored in different ways depending on the proposed use of the test scores. Curriculum bias may also vary based on the level of analysis.

Another important question is: How should we conceptualize and measure opportunity to learn? In this study opportunity to learn is used to represent the degree of content overlap between what is tested and taught (Husen, 1967). The measurement of opportunity to learn is problematic and has been discussed by Leinhardt and Seewald (1981), Leinhardt (1983) and Schmidt (1983). A complete treatment of this problem is beyond the scope of this paper. In this paper, opportunity to learn was obtained from teachers who examined each item in the test and reported whether or not the objective had been taught in their classrooms.

Given the potential biasing effects of lack of opportunity to learn the objectives measured on the test, what should be done? How

can we discover any systematic "error" in this situation? How can the potential curriculum bias due to lack of overlap between what is tested and taught be minimized? One approach to the problem of curriculum bias is to view the error due to lack of opportunity to learn as a random source of error. In this case, a robust estimator of student achievement can be used to minimize the error in a manner analogous to the way in which error due to random guessing and carelessness are minimized. A general class of robust estimators for ability or achievement can be obtained in a manner similar to weighted least squares. For example, Mislevy and Bock (1982) have proposed a robust estimator based on Tukey's biweight. They justify the use of the biweight estimator on the following basis,

> It seems reasonable to pay less attention to a subject's responses to items which are extremely hard or extremely easy for him, since they are at once less informative and more prone to measurement disturbances. . . . we attempt to utilize each observation in proportion to its apparent value.
>
> (Mislevy and Bock, 1982, p. 728)

One potential problem with the use of a robust estimator is that curriculum bias may not be a random source of error. A second approach is to retain the idea of weights and to develop a set of a priori weights based on judgments about the relative value and importance of the educational objectives represented by the test items. In the estimation of student achievement, an indicator of opportunity to learn can be used to derive a suitable set of weights which can be used to develop an alternative method of scoring the test. These curriculum weights can be based on an external judgment of the value of the items rather than on an internally derived set of

weights based on a robust estimator, such as the biweight. A set of curriculum-based estimates of student achievement using a priori judgmental weights (dichotomous or continuous) based on the students' opportunity to learn can be obtained in conjunction with the standard maximum likelihood estimates.

Are judgmental item weights a good idea? In general, the use of judgmental weights has been problematic. A major problem is the accuracy of these weights. In spite of the recognized problems, the use of weights to yield improved estimation procedures within the context of least squares has a long history. As pointed out by Mosteller and Tukey (1977),

> In surveying and in astronomy, where least squares originated, investigators long ago recognized that some observations are "better" or "stronger" than others and took appropriate action [emphasis added]. This action often assigned differing weights to different observations, either for objective reasons or as a matter of judgment. Thus the history of weighted least squares is almost as extensive as that of ordinary least squares.
> (Mosteller and Tukey, 1977, p. 346)

The curriculum-based estimates of student achievement proposed in this study using opportunity to learn represented by a suitable weighting function is an attempt to take "appropriate action" in situations where the content overlap between what is tested and what is included in the curriculum may introduce a significant curriculum bias. The rationale for using opportunity to learn to obtain a curriculum-based estimates is based on the idea that the estimates of student achievement should be based on the objectives that the students have had the opportunity to learn in the school curriculum.

An important idea is that the students' responses to items which

were not included in the curriculum are more likely to contain error, and we want to develop a set of weights to minimize this error. The issue of fairness is also important--students should only be tested on items which they have had a "fair" chance to learn. These curriculum-based weights can be dichotomous which would reflect the view that the items with weights of zero are of no value, while items with weights of one are of high value. This is explicitly what occurs in the design and development of customized tests. Values for the weights between zero and one can also be used to reflect the relative value of the items in more detail.

Purpose

The purpose of this study is to describe and illustrate an approach which can be used to obtain curriculum-based estimates of student achievement by including concomitant information about the school curriculum directly in an item response model. These curriculum-based estimates of student achievement can be obtained through a very simple modification of the maximum likelihood equations using a set of item weights designed to reflect the potential effects of the school curriculum. These judgmental weights can be derived from a variety of sources. The use curriculum-based estimates was illustrated with a set of mathematics achievement items from the Second International Mathematics Study.

Curriculum-Based Estimates of Student Achievement

The likelihood function for obtaining maximum likelihood estimates of student achievement, θ, can be expressed as

$$L(\underline{x}|\theta) = \prod_{i=1}^{n} P_i(\theta)^{x_i} \; [1 - P_i(\theta)]^{1 - x_i} \qquad (1)$$

where $\underline{x}$ is a vector of n dichotomous item responses, x_i is the response of the student to item i (Ø = failure, 1 = success), n is the number of items on the test and $P_i(\theta)$ represents the probability of the student succeeding on item i based on a suitable item response model.

If the item parameters are known for the n items, and if we assume that the responses are independent, given θ, then Equation 1 represents the probability of observing a particular vector of responses. The maximum likelihood estimate of the student's achievement, θ, is the value which maximizes Equation 1. Maximum likelihood estimators have the following general form:

$$\sum_{i=1}^{n} w_i(\theta_i) \; [\, x_i - P_i] = \emptyset \qquad (2)$$

where $w_i(\theta)$ represents the appropriate weighting function for item i which is dependent on the particular item response model selected. (See Wainer and Thissen (1985) for a description of several estimators based on different weighting functions).

In practice, the log of Equation 1 can be maximized using a suitable numerical method for solving implicit non-linear equations of this form, such as Newton-Raphson. In the case of the two-parameter item response model (item difficulty and discrimination parameters), the form of the Newton-Raphson iterations which can be used to obtain the maximum likelihood estimates of student achievement is as follows:

$$\theta_{ML} = \theta^{k+1} = \theta^{k} - \frac{\sum_{i=1}^{n} [x_i - P_i(\theta^k)]\, a_i}{-\sum_{i=1}^{n} [P_i(\theta^k)\ (1 - P_i(\theta^k))]\, a_i^2} \qquad (3)$$

where θ^k is the initial estimate, θ^{k+1} is an updated estimate, and a_i is the discrimination parameter for item i. These iterations can be continued until an appropriate stopping criterion is reached.

In order to obtain the curriculum-based estimates of student achievement, Equation 3 can be modified to explicitly contain a set of weights, w_i, which reflect the relative emphasis on the item objective in the curriculum. The curriculum-based estimates can be obtained as follows:

$$\theta_{CB} = \theta^{k+1} = \theta^{k} - \frac{\sum_{i=1}^{n} w_i\ [x_i - P_i(\theta^k)]\, a_i}{-\sum_{i=1}^{n} w_i\ [P_i(\theta^k)\ (1 - P_i(\theta^k))]\, a_i^2} \qquad (4)$$

where the weights, w_i, can be dichotomous (1 = high opportunity to learn, 0 = low opportunity to learn) or continuous weights between 0 and 1 to reflect in detail the relative value of the items. This modification follows the suggestion made by Mislevy and Bock (1982) for obtaining biweight estimates of ability. The major difference is that the weights used to obtain the curriculum-based estimates of student achievement are obtained a priori on the basis judgments about

the relative value of each item objective in the class curriculum, while the weights used to obtain the biweight estimates are internally derived.

A large sample standard error for the curriculum-based estimates can be obtained as follows

$$SE(\Theta_{CB}) = \sum_{i=1}^{n} [\, w_i \, [P_i(\Theta^k) \, (1 - P_i(\Theta))] \, a_i{}^2]^{-1/2} \qquad (5)$$

after obtaining a converged estimate of student achievement, Θ.

It is clear that the maximum likelihood estimates of student achievement can be obtained from Equation 4 by setting the curriculum-based weights, w_i equal to one for all of the items. This reflects the idea that all of the items are of equal value in determining student achievement, while with the curriculum-based estimates of student achievement a weighted average is obtained based on some evaluation of the relative value of each item objective, such as teacher judgments of opportunity to learn.

Example

In order to illustrate how curriculum-based estimates of student achievement can be used, a small example is presented in Table 1.

Insert Table 1 about here

This table was created by starting with 10 students with known achievement values, Θ, ranging from -2.0 to 2.0. Students 1 and 2 have the same generating achievement value of -2.0, students 3 and 4 have the same generating value of -1.0 and so on. A 22 item test with

item difficulties ranging from -2.94 to 2.94 was used to simulate the item responses for these 10 students.

As pointed out earlier, curriculum bias can be viewed as the difference between the "true" achievement level of the student and the obtained estimates. In practice, the "true" achievement of the students are of course not known and curriculum bias can be operationally defined as the difference between the maximum likelihood and curriculum-based estimates. This difference can be positive or negative. The maximum likelihood estimates may be larger than the curriculum-based estimates if the students have inflated scores due to guessing. On the other hand, the curriculum-based estimates might be larger, if the students fail on items which they have not had an opportunity to learn. This "penalty" may lead to a decrease in the probability of a student succeeding on an item when he or she has not had an opportunity to learn the objectives measured by the item.

In order to illustrate the method, curriculum bias will be viewed as a penalty and the potential effects which may result from a low opportunity to learn. The 22 item test was divided into two parts with equal item difficulties in each half. The first eleven items were classifed as having <u>low</u> curriculum dependence, while the second eleven items were classified as having <u>high</u> curriculum dependence.

The idea of curriculum dependence simply means that if the students have not had the opportunity to learn objectives which are highly curriculum dependent, then the probability of succeeding on these items will be decreased. The probability of succeeding on items measuring objectives which have low curriculum dependence will not be

affected by whether or not the student has had an opportunity to learn the objectives. For example, some mathematics objectives would be highly dependent on being learned in the curriculum, since the students would have fewer opportunities to learn the these objectives outside of school. Items measuring reading comprehension on the other hand may be less dependent on the school curriculum because of the many opportunities to learn to read outside of the formal school currciculum. This concept of curriculum dependence reflects the major reason why opportunity to learn is a significant variable in explaining student achievement.

Each of the students has been classified as having either a high or low opportunity to learn the objectives represented by the 22 item test. Student 1 has a low achievement level and has had a high opportunity to learn the objectives measured by the 22 item test. He succeeds on items 1 and 2, as well as items 12 and 13 as expected given the generating acheivement value and the difficulties of these items. Student 2 has <u>not</u> had the opportunity to learn the objectives covered on the test, and she is able to succeed on items 1 and 2 as expected, however she is not able to succeed on items 12 and 13. She is being penalized because items 12 to 22 are highly curriculum dependent. Since she has not had an opportunity to learn these objectives, she fails on items that she would be expected to succeed on <u>if</u> she had an opportunity to learn these items.

Since the data has been generated and the true achievement levels for these students are known, the impact on three estimates of student

achievement can be examined. Clearly, Student 2 has had her achievement level underestimated. In terms of raw scores, she has a score of 2 as compared to the expected raw score of 4. The maximum likelihood estimates also underestimate the generating achievement value. If the opportunity to learn these objectives is taken into account through the use of curriculum-based estimates of achievement based on the weights given in Table 1, then the curriculum-based estimate is closer to the generating value. The standard errors for the curriculum-based estimates are larger which reflects the loss of information due to the deletion of test items through the use of dichotomous weights. The differences between the maximum likelihood and curriculum-based estimates can be used as a indication of curriculum bias. For Student 2, her achievement is underestimated by -1.04. The curriculum biases for the other students are also shown in Table 1 and range from -.40 to -1.17. These curriculum biases are shown graphically in Figure 1.

Insert Figure 1 about here

One question that can be raised at this point is: Do we want to consider Student 1 and Student 2 as having the same level of achievement? Clearly, Student 2 has not mastered the objectives measured by items 12 and 13, however is it "fair" to ignore the additional data which is available on her opportunity to learn these objectives? There is no empirical way to resolve this question, and perhaps the best approach might be to simply calculate both estimates

of achievement and see if there is a significant curriculum bias. The use of raw scores does not allow this option, but if the application of item response theory is appropriate, then the computation of both estimates becomes practicable.

Application

Sample

The Second International Mathematics Study (SIMS) is a comprehensive study of the teaching and learning of mathematics conducted in about two dozen countries during the 1981-82 school year. In the United States, students and in teachers in over 500 eighth grade and twelfth grade classrooms were studied. A complete description of the study is provided in several reports (Crosswhite, et al., 1985; McKnight, et al. 1987).

The analyses presented in this paper are based on the responses of eighth grade students in the United States who were enrolled in classrooms that teachers classified as "typical". Students were included in the sample if they had complete pretest and posttest information on the 40 item mathematics core test, and if information on student opportunity to learn was available for their classrooms. Only the posttest responses of the students were used in this study. A total of 165 classrooms and 2,606 students were included in the final sample. The reports cited above should be consulted for a detailed description of the sampling procedures used in SIMS.

Procedure

A set of 16 arithmetic items from the 40 item core test which was administered to all students was selected to illustrate the utility of

the curriculum-based estimates of student achievement. These 16 arithmetic items were calibrated using a one-parameter Rasch model (Rasch, 1960) based on the total sample of 2,606 students. The texts of the arithmetic items are given in Chang and Ruzicka (1985).

Once the arithmetic items were calibrated, the maximum likelihood and curriculum-based weights of student achievement were obtained through a computer program written with PROC MATRIX (SAS, 1982). The curriculum-based estimates were obtained on the basis of a teacher's response to the following question: During this school year did you teach or review the mathematics necessary to answer this item correctly? Students in classrooms where teachers responded yes to this question were coded with a curriculum-based weight of 1, while a no response was coded as a 0.

In order to illustrate the potential advantages and disadvantages of these two estimators, a subset of students who had the opportunity to learn 50 to 75 percent of the items was identified and used in the analyses. Each classroom had its own unique curriculum-based weights based on the teachers' reports of opportunity to learn.

Curriculum bias, CB, can be defined as follows:

$$CB = \theta_{ML} - \theta_{CB} \tag{6}$$

and a standardized index of curriculum bias, SCB, which takes into account the standard errors of the two estimates can be defined as:

$$SCB = \frac{[\theta_{ML} - \theta_{CB}]}{[(SE(\theta_{ML})^2 + SE(\theta_{CB})^2]^{1/2}} \tag{7}$$

If this index is greater than 2.0, then there is some rough indication that the curriculum bias is statistically significant. Since the two estimators are based on overlapping item data, a more rigorous test would have to take this dependence into account.

Results

The p-values, preliminary scale values (item difficulties), standard errors and teacher reports of student opportunity to learn for the 16 arithmetic items are presented in Table 1. The items range

---

Insert Table 2 about here

---

from item 6 with approximately 4 percent of the 2,606 students succeeding on this item to item 8 with slightly more than 60 percent of the students able to answer correctly. Teacher reports of student opportunity to learn are generally quite high. When the dichotomous curriculum-based weights are summed for each classroom (N = 165) to obtain a total opportunity to learn score for each classroom, 33 percent of the teachers report that students have had the opportunity to learn all 16 items. Seventy-four percent of the teachers report that students in their classrooms had the opportunity to learn 14 or more of the objectives represented by these items. In general the match between the items and teacher coverage is quite good.

There were 510 students in classrooms where the opportunity to learn ranged from 50 to 75 percent based on teacher reports (25 classrooms). A plot of the maximum likelihood and curriculum-based estimates for these students is given in Figure 2. Usable

---

Insert Figure 2 about here

---

estimates were obtainable for 490 students. The other 20 students had item response patterns which did not lead to converged estimates because of all wrong responses or other troublesome patterns; none of the students succeeded on all 16 arithmetic items. A correlation of .89 was found between the maximum likelihood and curriculum-based estimates.

A plot of the relationship between curriculum bias (maximum likelihood minus curriculum-based estimates) and the maximum likelihood estimates is presented in Figure 3.

Insert Figure 3 about here

As would be expected, there is considerable variation in the amount of curriculum bias. The largest underestimate of student achievement was -1.61. For 17 percent of the students (N = 82), the difference between the maximum likelihood and curriculum-based estimates was underestimated by at least .5 logits. The greatest overestimate was 1.34 and approximately 14 percent of students (N = 69) had the achievement overestimated by at least .5 logits.

Although it might be argued that these differences are large enough to be considered of substantive significance, the question of whether these differences are statistically significant is important. Some indication of the significance of the differences can be obtained through the use of a standardized curriculum bias index described in Equation 7. The relationship between the standardized index of

curriculum bias and the maximum likelihood estimates is presented in Figure 4.

Insert Figure 4 about here

Any values greater than 2.0 on this index can be viewed as reflecting a statistically significant difference. None of the differences are significant based on this criterion. This result is not surprising given the large standard errors of the two estimators given the small number of items, and also the good match between student opportunity to learn and the objectives measured by the test items.

Another way to summarize the data is to form score groups on the basis of the maximum likelihood estimates, and to compute summary statistics for the maximum likelihood and curriculum-based estimates within these score groups. The results by score group are reported in Table 3.

Insert Table 3 about here

In score group 2, there is some indication that the achievment level for these 17 students is underestimated, but the lack of variation in the maximum likelihood estimates suggests that these students should be examined more closely. When the standard errors are taken into account, these differences are not larger than would be expected by chance. The data do <u>not</u> provide any strong evidence of a systematic curriculum bias in the other score groups. In score groups 3 to 6, the average differences between the maximum likelihood and curriculum-based estimates are well within the range of differences

which would be expected given the standard errors. In the other extreme group, there were only 2 students and the difference between the two estimators does not warrant any interpretation.

Discussion

The curriculum-based estimates proposed in this paper offer an alternative approach for examining and adjusting student achievement estimates in order to take into account whether or not the student has had an opportunity to learn the objectives included in the test. When curriculum-based weights are used, the responses of each student are weighted in order to minimize the potential effects of curriculum bias. Curriculum bias is viewed as source of error which can have a significant impact on the reliability and validity of the student responses.

The extent to which the use of curriculum-based weights will indeed reduce curriculum bias and lead to more accurate estimates of student achievement is an empirical question. In the small example which was constructed to illustrate the use of curriculum-based estimates, the use of these weights seemed plausible. When the curriculum-based estimates were used with the SIMS arithmetic test items, the data suggest that there may be some curriculum bias but an approximate statistical test of the differences did not reveal any statistically significant differences. The absolute value of the curriculum bias in some cases may be of substantive significance, however more experience with the curriculum-based estimates will be required before any strong conclusions can be reached.

Since the major purpose of this paper is to describe and illustrate the method, no strong substantive conclusions should be drawn based on the small number of items (16) included in the application. One problem which must be addressed is how to calibrate the item bank. This problem is not unique to the proposed curriculum-based estimates, but is crucial in many applications of item response theory. In the application, the test items were calibrated based on the total group. This total group included the subgroup which was used in the subsequent analyses. Perhaps the lack of differences between the maximum likelihood and curriculum-based estimates is simply due to the effects of opportunity to learn being averaged out during the item calibration process.

Another significant problem which must be addressed before the curriculum-based estimates can be used is related to the measurement of student opportunity to learn. Are teacher responses a reliable and valid source of information on school curriculum? Better indicators of student opportunity to learn might be developed by using other sources of information, such as observations of classrooms, analyses of textbooks and even asking the students about their opportunity to learn.

In spite of the potential difficulties, the results of this study suggest that the use of curriculum-based weights to obtain a customized test for each student is practicable. Ideally, students should only have to respond to test items that are appropriate for them. The identification of "appropriate" items should include a consideration of content and whether or not the student has had an

opportunity to learn the objectives measured on the test. In some cases, the inferences and decisions for which the tests will be used may require a set of test items which match local objectives and in other cases the match between the curriculum and test may not matter. The curriculum-based estimates can be used in situations where potential curriculum bias is of concern and both estimators used to determine the impact on student achievement estimates.

Deficiencies in the test development and item selection process as well as practical problems may prevent the complete tailoring of test items which are appropriate for every student. When this is the case and a suitably calibrated item bank is available, then the curriculum-based estimates described in this paper offer an approach which can be used to obtain adjusted estimates of student achievment which reflect opportunity to learn.

References

Anderson, L.W. (1985). Opportunity to learn. In Husen, T. & Postlethwaite, T.N. (Eds). The International Encyclopedia of Education, Oxford: Pergamon Press.

Borg, W.R. (1979). Teacher coverage of academic content and pupil achievement. Journal of Educational Psychology, 71, 635-645.

Chang, L. & Ruzicka, J. (1985). Second International Mathematics Study, Technical Report 1: Item level achievement data, Eighth and Twelfth Grades. Champaign, Il: Stipes Publishing Co.

Crosswhite, F.J., Dossey, J.A., Swafford, J.O., McKnight, C.C. & Cooney, T.J. (1985). Second International Mathematics Study: Summary Report for the United States. Champaign, Ill.: Stipes Publishing Company.

Hambleton, R.K. & Swaminathan, H. (1985). Item response theory: Principles and applications. Boston, MA: Kluwer-Nijhoff Publishing.

Husen, T. (Ed.) (1967). International Study of Achievement in Mathematics: A Comparison of Twelve Countries. New York: John Wiley & Sons.

Jenkins, J.R. & Pany, D. (1978). Curriculum biases in reading achievement tests. Journal of Reading Behavior, 10, 345-347.

Leinhardt, G. (1983). Overlap: Testing whether it is taught. In Madaus, G. F. (Ed.), The Courts, Validity, and Minimum Competency Testing. Boston: Kluwer-Nijhoff Publishing.

Leinhardt, G. & Seewald, A.M. (1981). Overlap: What's tested, what's taught. Journal of Educational Measurement, 18, 85-96.

McKnight, C.C., Crosswhite, F.J., Dossey, J.A., Kifer, E., Swafford, J.O., Travers, K.J., & Cooney, T.J. (1987). The underachieving curriculum: Assessing U.S. school mathematics from an international perspective. Champaign, IL: Stipes Publishing Co.

Mehrens, W.A. & Phillips, S.E. (1986). Detecting impacts of curricular differences in achievement test data. Journal of Educational Measurement, 23, 185-196.

Miller, M.D. (1986). Time allocation and patterns of item response. Journal of Educational Measurement, 23, 147-156.

Mislevy, R.J. & Bock, R.D. (1982). Biweight estimates of latent ability. Educational and Psychological Measurement, 42, 725-737.

Mosteller, F. & Tukey, J.W. (1977). Data analysis and regression. Reading, MA: Addision-Wesley Publishing Company.

Pelgrum, W.J., Eggen, T.J.H.M. & Plomp, Tj. (1986). The implemented and attained mathematics curriculum: A comparison of eighteen countries. The Netherlands: Department of Education, Twente University of Technology.

Rasch, G. (1960). Probabilistic models for some intelligence tests. Copenhagen: Dansmarks Paedagogiske Institut. (Reprinted 1980, Chicago: The University of Chicago Press.

SAS Institute (1982). SAS Users Guide: Statistics, 1982 Edition. Cary, NC: Author.

Schmidt, W.H. (1983). Content biases in achievement tests. Journal of Educational Measurement, 20, 165-178.

Wainer, H. & Thissen, D. (1985). Estimating ability with the wrong model. Brooks Air Force Base, Texas: Air Force Human Resources Laboratory.

Weiss, D.J. (1982). Improving measurement quality and efficiency with adaptive testing. Applied Psychological Measurement, 10, 473-492.

Table 1

Maximum likelihood and curriculum-based estimates of student achievement for several hypothetical item response patterns

Student	θ	OTL	Item Response Vectors Curriculum-Dependence Low	High	Achievement Estimates RS	ML	CB	Bias
1	-2.0	High	11000000000	11000000000	4	-2.28 (.65)		
2	-2.0	Low	11000000000	[00000000000]	2	-3.32 (.82)	-2.28 (.92)	-1.04
3	-1.0	High	11100000000	11100000000	6	-1.52 (.59)		
4	-1.0	Low	11100000000	[10000000000]	4	-2.28 (.65)	-1.52 (.83)	-.76
5	0.0	High	11111100000	11111100000	12	.29 (.54)		
6	0.0	Low	11111100000	[11000000000]	8	-.88 (.55)	.29 (.76)	-1.17
7	1.0	High	11111111000	11111111000	16	1.52 (.58)		
8	1.0	Low	11111111000	[11111000000]	13	.58 (.54)	1.52 (.83)	-.94
9	2.0	High	11111111100	11111111100	18	2.28 (.65)		
10	2.0	Low	11111111100	[11111111000]	17	1.88 (.61)	2.28 (.92)	-.40

Item Difficulties:

-2.94, -2.20, -1.39, -.85, -.41, .0, .41, .85, 1.39, 2.20, 2.94
-2.94, -2.20, -1.39, -.85, -.41, .0, .41, .85, 1.39, 2.20, 2.94

Weights: ML 11111111111 11111111111
CB 11111111111 00000000000

Table 2

Preliminary Calibration of 16 Item Core Arithmetic Test and Teacher Reports of Student Opportunity to Learn

Item	SIMS Code	P-Value	Scale Value	Standard Error	Opportunity to Learn
1	189	.567	-1.205	.052	.923
2	076	.268	.461	.058	.787
3	079	.381	-.220	.053	.878
4	181	.092	2.137	.089	.760
5	045	.265	.479	.058	.788
6	109	.039	3.330	.132	.884
7	190	.505	-.882	.051	.847
8	005	.612	-1.447	.053	.932
9	075	.547	-1.101	.051	.716
10	003	.556	-1.150	.052	.941
11	140	.143	1.489	.073	.941
12	008	.478	-.737	.051	.874
13	179	.324	.108	.055	.917
14	009	.502	-.862	.051	.857
15	043	.304	.232	.056	.654
16	046	.458	-.632	.051	.929

Note. Calibration is based on students in classrooms classifed as typical by the teachers (N = 2,606), teacher reports of student opportunity to learn is also based on typical classrooms (N = 165). See Chang & Ruzicka (1985) for item texts.

Table 3

Summary of SIMS Arithmetic Data by Score Groups

Score Group	Estimator	N	Mean	Standard Error of the Mean
2	ML	17	-3.20	.00
(-3.5 to -2.5)	CB	17	-2.72	.04
	Mean Difference		-.48	
3	ML	123	-2.03	.02
(-2.4 to -1.5)	CB	123	-1.99	.05
	Mean Difference		-.04	
4	ML	197	-1.11	.02
(-1.4 to -.5)	CB	197	-1.08	.04
	Mean Difference		-.03	
5	ML	108	-.17	.02
(-.4 to .5)	CB	108	-.13	.06
	Mean Difference		-.04	
6	ML	43	.85	.04
(.6 to 1.5)	CB	43	.78	.08
	Mean Difference		.07	
7	ML	2	1.89	.00
(1.6 to 2.5)	CB	2	1.49	.12
	Mean Difference		.40	

Note. The score groups were formed on the basis of the maximum likelihood estimates. The range used for each score group is shown in parentheses.

Figure 1. Expected Relationship Between Maximum Likelihood and Curriculum-Based Estimates of Student Achievement

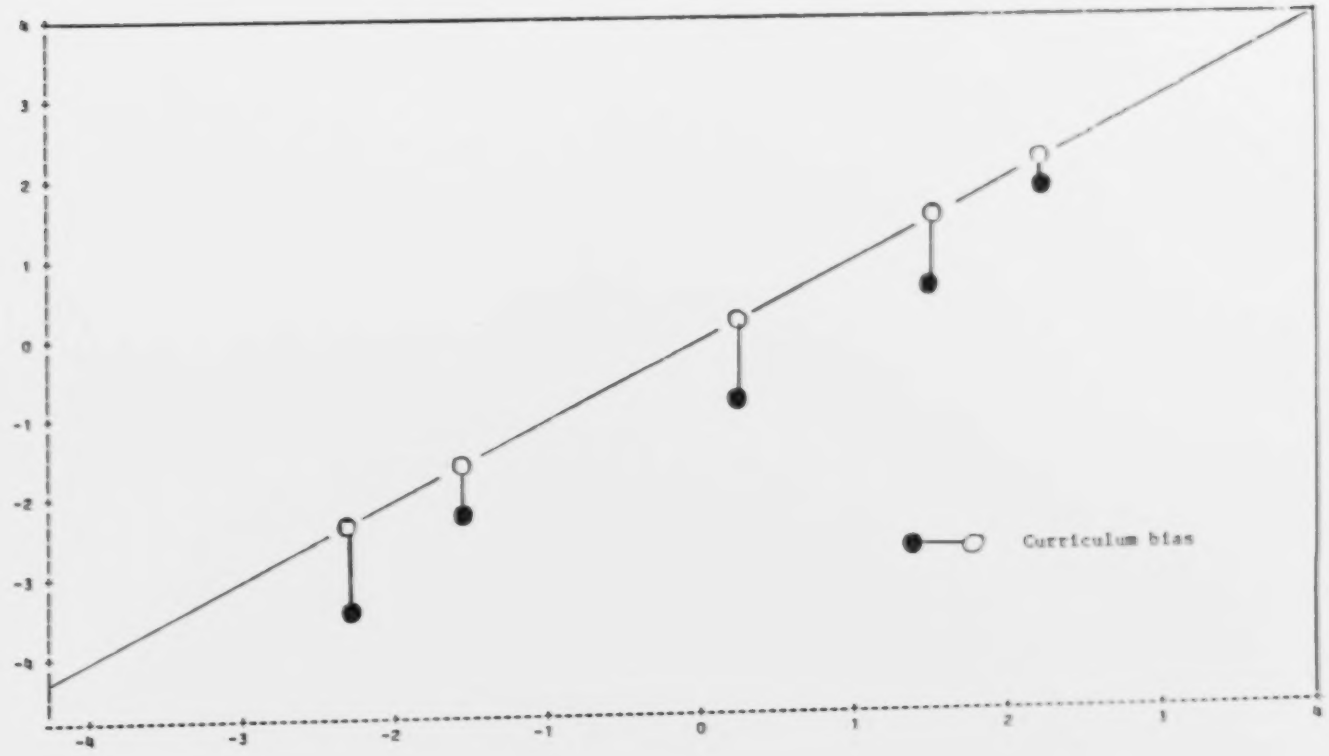

Figure 2. Relationship Between Maximum Likelihood and Curriculum-Based Estimates of Student Achievement for 16 Item Arithmetic Test (N = 490)

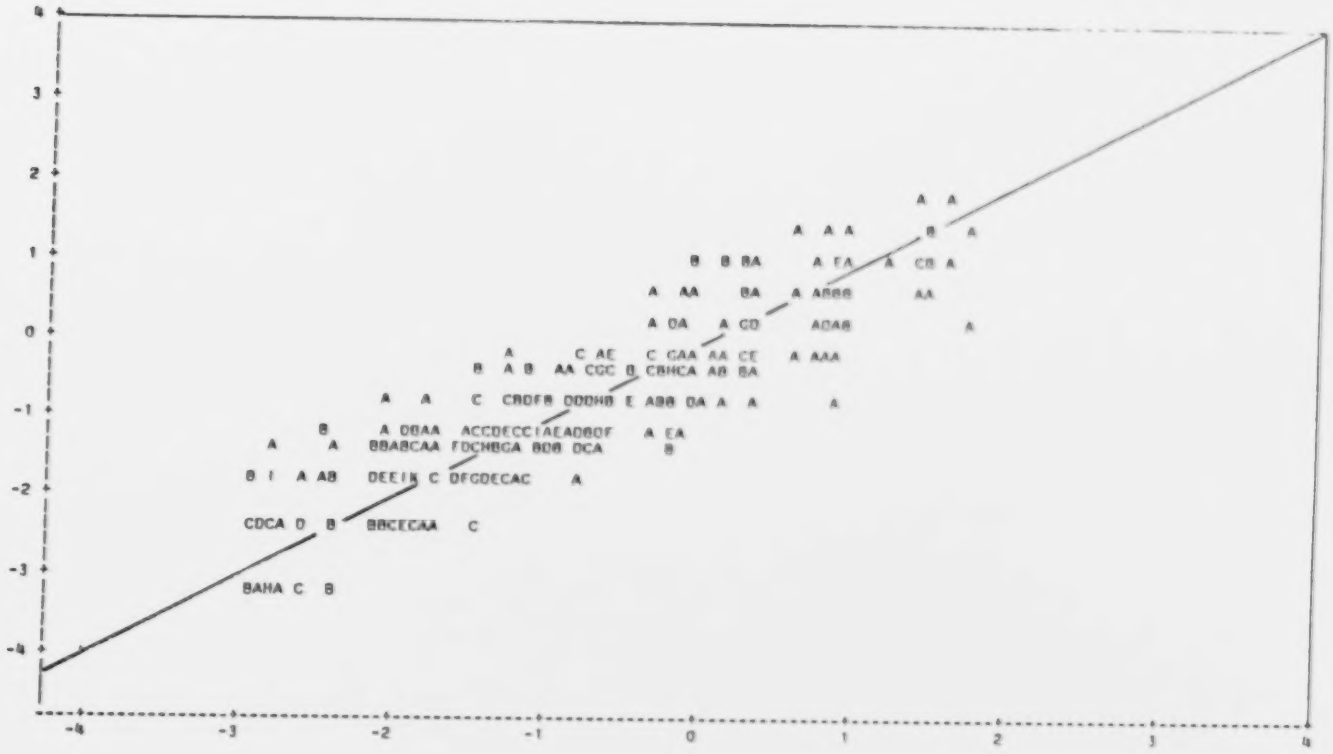

Figure 3. Relationship Between Curriculum Bias (Maximum Likelihood Minus Curriculum-Based Estimates) and Maximum Likelihood Estimates

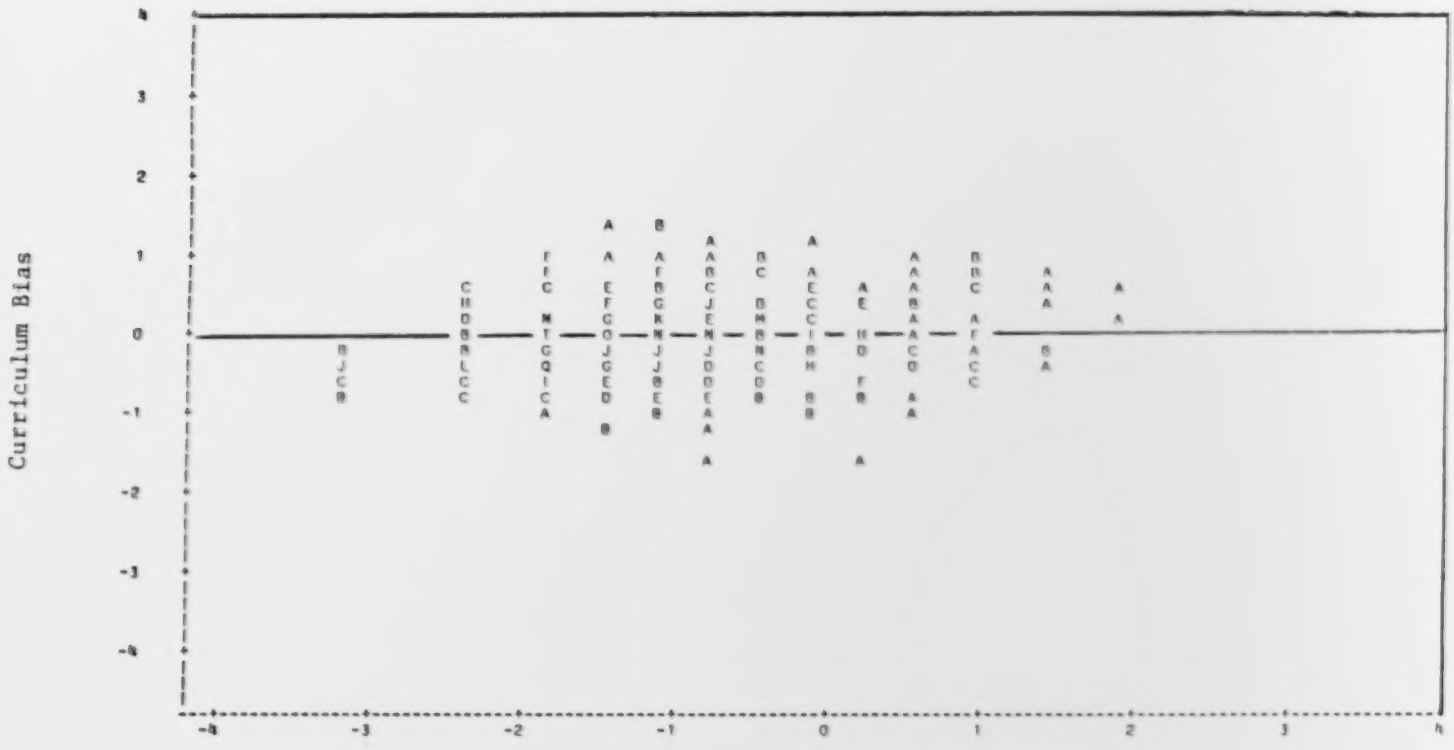

Figure 4. Relationship Between Standardized Curriculum Bias and the Maximum Likelihood Estimates